

# THE SEILS DATASET: SYMBOLICALLY ENCODED SCORES IN MODERN-EARLY NOTATION FOR COMPUTATIONAL MUSICOLOGY

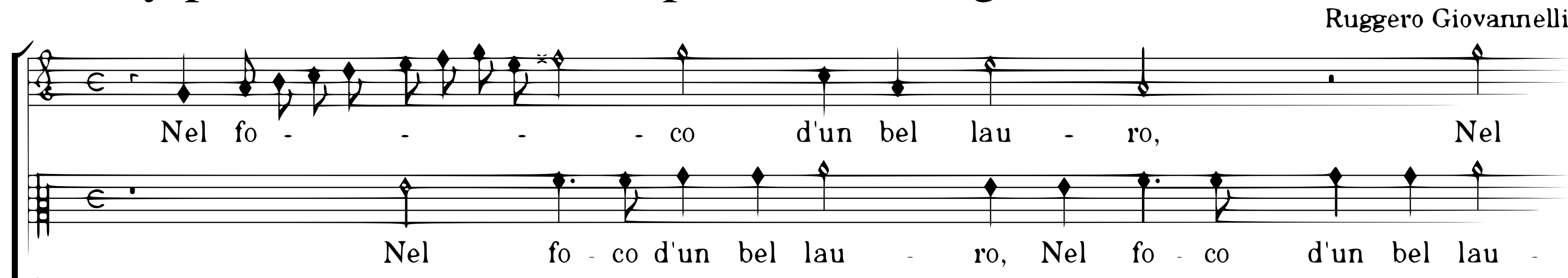
Emilia Parada-Cabaleiro<sup>1,2</sup>, Anton Batliner<sup>1,2</sup>, Alice Baird<sup>1,2</sup>, Björn Schuller<sup>1,2,3</sup>

<sup>1</sup>Chair of Complex and Intelligent Systems, University of Passau, Germany

<sup>2</sup>Chair of Embedded Intelligence for Health Care and Wellbeing, Augsburg University, Germany

<sup>3</sup>GLAM – Group on Language, Audio & Music, Imperial College London, U.K.

The symbolic codification of Renaissance music is restricted, mainly preserved scanned copies of the original scores



## The SEILS Dataset

(Symbolically encoded *Il Lauro Secco*)

30 symbolic transcriptions in Early/Modern notation of the inedited Renaissance Anthology of Italian Madrigals “*Il Lauro Secco*”

The Italian Madrigal of the 16<sup>th</sup> century is a composition for small vocal ensembles characterised by the lyrical mimicry throughout the music

### *Il Lauro Secco* Anthology

- 30 Italian Madrigals for 5 unaccompanied voices
- Encoded in *White mensural* notation
- High musical-linguistic coherence
- Written by 30 different composers
- 2 editions: Baldini (1582) and Gardano (1596)

## Symbolic score codification in Modern and Early notation\*\*

### Musical Criteria

- Vertical alignment of voices (original in individual sheets)
- Accidentals and “no” barlines (according original source)

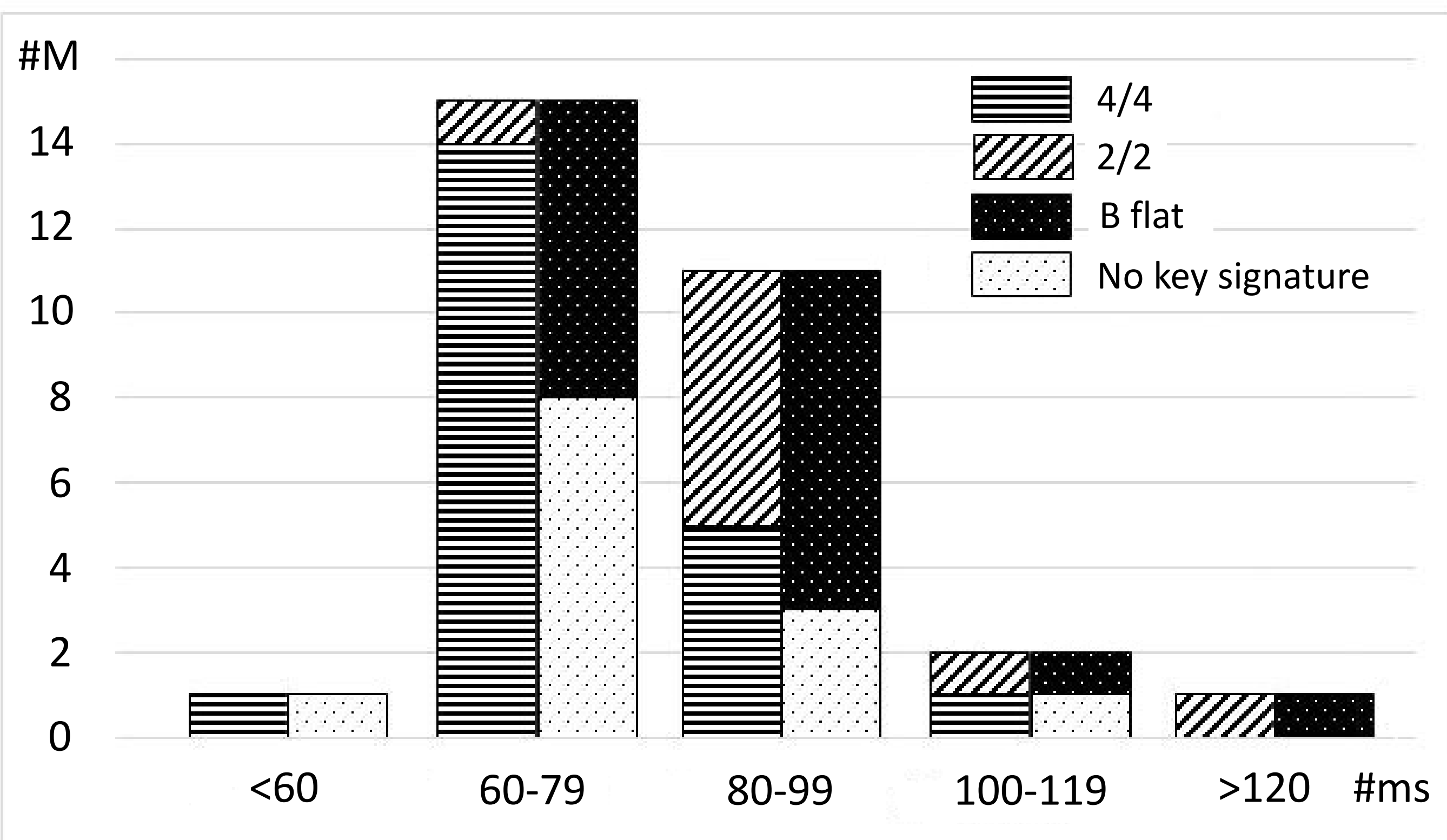
### Linguistic Criteria

- The arbitrary use of ‘v’ and ‘u’ have been normalized
- Abbreviations, as e.g. ~ or *ij* have fully spelled



\*\*The transcription refers to the source printed in 1582

## The SEILS Dataset Statistic Evaluation



#M: number of Madrigals

#ms: number of measures

	16th	8th	4th.	2 <sup>nd</sup>	breve	acc
Belli	0	42	17	251	4	63
Eremita	0	127	61	167	7	34
Fiorino	8	62	19	295	0	25
Luzzaschi	0	65	15	348	5	11
Macque	0	265	48	170	1	34
Massaino	0	173	40	248	12	36
Perue	2	35	11	168	0	21
Spontone	2	73	28	269	2	7
Striglio	0	252	85	271	2	29
Zoilo	0	27	2	187	2	25
30M*	60	3222	958	7399	117	817
Mean (30M)	2	107.4	31.9	246.6	3.0	27.3
Sd(30M)	2.4	66.8	19.7	52.8	3.6	14.4

\*\*30M refers statistics for the whole dataset

## Digital formats evaluation

4 digital formats: Lilypond, Music XML, MIDI, and Finale  
Early/Modern encoded in Lilypond – Modern in others formats.

### MUSIC XML

```

1 <note default-x="121">
2 <pitch>
3 <step>B</step>
4 <alter>-1</alter>
5 <octave>4</octave>
6 </pitch>
7 <duration>8</duration>
8 <voice>1</voice>
9 <type>whole</type>
10 <lyric default-y="-80" number="1">
11 <syllabic>begin</syllabic>
12 <text>Men</text>
13 </lyric>
14 </note>
    
```

### LILYPOND

```

1 \key f \major
2 \time 4/4
3 \autoBeamOff
4 bes'1 |% 1
5 a4 bes4. bes8 c4 |% 2
6 d bes a8 g f e |% 3
7 d4 bes' a2 |% 4
8 a bes |% 5
9 c4. c8 c4 d |% 6
    
```

In 14 lines of Music XML,  
1 note is encoded  
In 9 lines of Lilypond,  
20 notes are encoded

## CONCLUSIONS

SEILS Dataset specifically emphasises the **early/modern** transcriptions in **Lilypond**, a symbolic language that allows for music engraving, offering a clearer understanding than more symbolic musical formats. This has great application potential for computational musicology in real environments, such as musical libraries or pedagogy.